Bridging the gap: leveraging data science to equip domain experts with the tools to address challenges in maternal, newborn, and child health

Check for updates

Girmaw Abebe Tadesse[1] ✉, William Ogallo[2], Celia Cintas[3], Skyler Speakman[3], Aisha Walcott-Bryant[2] & Charity Wayua[3]

The United Nations Sustainable Development Goals (SDGs) advocate for reducing preventable Maternal, Newborn, and Child Health (MNCH) deaths and complications. However, many low- and middle-income countries remain disproportionately affected by high rates of poor MNCH outcomes. Progress towards the 2030 sustainable development targets for MNCH remains stagnated and uneven within and across countries, particularly in sub-Saharan Africa. The current scenario is exacerbated by a multitude of factors, including the COVID-19 pandemic's impact on essential services and food access, as well as conflict, economic shocks, and climate change. Traditional approaches to improve MNCH outcomes have been bifurcated. On one side, domain experts lean heavily on expert-driven analyses, often bypassing the advantages of data-driven methodologies such as machine learning. Conversely, computing researchers often employ complex models without integrating essential domain knowledge, leading to solutions that might not be pragmatically applicable or insightful to the community. In addition, low- and middle-income countries are often either data-scarce or with data that is not readily structured, curated, or digitized in an easily consumable way for data visualization and analytics, necessitating non-traditional approaches, data-driven analyses, and insight generation. In this perspective, we provide a framework and examples that bridge the divide by detailing our collaborative efforts between domain experts and machine learning researchers. This synergy aims to extract actionable insights, leveraging the strengths of both spheres. Our data-driven techniques are showcased through the following five applications: (1) Understanding the limitation of MNCH data via automated quality assessment; (2) Leveraging data sources that are available in silos for more informed insight extraction and decision-making; (3) Identifying heterogeneous effects of MNCH interventions for broader understanding of the impact of interventions; (4) Tracking temporal data distribution changes in MNCH trends; and (5) Improving the interpretability of "black box" machine learning models for MNCH domain experts. Our case studies emphasize the impactful outcomes possible through interdisciplinary collaboration. We advocate for this joint collaborative research approach, believing it can accelerate the extraction of actionable insights at scale. Ultimately, this will catalyse data-driven interventions and contribute towards achieving SDG targets related to MNCH.

[1]Microsoft AI for Good Research Lab, Nairobi, Kenya. [2]Google Research, Nairobi, Kenya. [3]IBM Research, Nairobi, Kenya. ✉e-mail: gtadesse@microsoft.com

The third agenda of the United Nation's (UN's) Sustainable Development Goals (SDGs) stresses the need to ensure healthy lives and well-being at all ages. Maternal, Newborn, and Child Health (MNCH) is a critical component of this agenda, particularly in low-and middle-income settings, where there is a significant gap in access and quality of health services[1]. In 2019, it was reported that about five million deaths of under-five children and 112 million maternal complications and deaths occurred globally[2]. Unfortunately, the situation was further exacerbated by the COVID-19 pandemic's impact on essential services and food access in addition to other challenges such as climate change, disasters, and wars, which make it challenging to achieve the apriori set SDG targets[3,4].

Existing approaches to address MNCH issues have been bifurcated. On one side, domain experts lean heavily on expert-driven analyses, such as determinants for child mortality[5–8]. However, these approaches often bypass the advantages of advanced data-driven methodologies such as machine learning (ML), which can augment the capabilities of domain experts or scale the analysis to large spatial and temporal dimensions. Moreover, such automated techniques can be used to extract actionable insights efficiently and contrast results across a large collection of health surveys from different geographical locations. Conversely, computing researchers often employ complex models without integrating essential domain knowledge, leading to solutions that might not be pragmatically applicable[9]. This is partially a result of concealed insights within health surveys, such as systematic deviations, which necessitate domain-expert knowledge to enhance the design of ML models[10]. Additionally, the synergistic collaboration between these two divergent approaches augments their respective capabilities, providing domain experts and/or policymakers with timely information. Such collaborations culminate in the creation of a streamlined, scalable, and efficient data analysis pipeline that (a) accommodates the ever-expanding volume of health data, (b) synergizes exploratory (often ML-driven) and confirmatory (often human-driven) methodologies, (c) uncovers concealed patterns supported by the most substantial evidence, and (d) safeguards against erroneous discoveries resulting from human bias or spurious outcomes sometimes produced by ML models.

In this perspective, we provide a framework and exemplar case studies that bridge the divide between MNCH domain experts and machine learning researchers by detailing our collaborative efforts and experiences. This synergy aims to extract actionable insights, leveraging the strengths of both spheres using the following five distinct use-cases:

- Understanding the limitation of MNCH data via data-driven and automated quality assessment
- Leveraging data sources that are available in silos for more informed insight extraction and decision making
- Identifying heterogeneous effects of MNCH interventions for wider impact understanding
- Tracking temporal data distribution changes in MNCH trends
- Improving the interpretability of "black box" machine learning models for MNCH domain experts

Our case studies emphasize the potent outcomes possible through interdisciplinary collaboration. We advocate for this joint approach, believing it can accelerate the extraction of insights on a grand scale. Ultimately, this will catalyse data-driven interventions and increase the probability of achieving SDG targets related to MNCH.

Understanding the limitation of MNCH data via data-driven and automated quality assessment

Data-driven analytical approaches are significantly dependent on the quality of the data used for analysis[11,12]. This is best encapsulated by the adage *garbage in, garbage out*, meaning that if the quality of the data used for analysis is substandard, the insights derived from the analysis are likely to be questionable[13,14]. The issue of data quality is particularly pertinent in the MNCH domain, where publicly accessible data is primarily collected through demographic surveys[15]. Health surveys such as the Demographic Health Survey (DHS)[16], Knowledge Integration (KI)[17], and Performance Monitoring

for Action 2020 (PMA)[18] can be analyzed to generate novel insights about MNCH[19]. However, a common pattern across these surveys is the *missingness* of key variables for a significant number of records[20]. In common data analysis practices, records with missing values are often discarded or imputed using mean or median values. However, it is key to uncover if there is a systematic pattern for the missingness, e.g., individuals with limited knowledge of the question or with privacy concerns may opt-out from answering a survey question. Furthermore, machine learning models often require well-balanced data across segments of the population of interest. Practically, that is hardly true as the data collected shows a varying degree of skewness towards a certain segment, e.g., due to proximity to data collection centers, cultural barriers to participate in data collection, or divergent prevalence of the outcome across regions. Thus, it is crucial to check the quality of health survey datasets, e.g., collection irregularities or skewed representations, before the survey is fed into ML models, which is also further supported by a growing trend of data-centric approaches for impactful ML solutions[21].

To this end, we aim to share how quality analysis is applied in evaluating health data using the BetterBirth[22] study as a use case. The BetterBirth Study was a matched-pair, cluster-randomized, controlled trial in 60 pairs of facilities across 24 districts of Uttar Pradesh, India, that studied the impact of the WHO's Safe Childbirth Checklist on adherence to evidence-based practices by birth attendants and a composite health outcome of perinatal and maternal deaths and serious complications[22–24]. Before investigating heterogeneous treatment effects of the intervention in the BetterBirth study, we first wanted to identify the subset of mothers who had the highest risk of neonatal deaths, i.e., *mothers who experienced the death of a new born within 28 days of delivery*, in both the Control and Intervention arms (see Table 1). The treatment arms in the BetterBirth study consisted of approximately 75,000 participants in the Control arm and 77,000 in the Intervention arm. The average rate of neonatal death was 3.15% in the Control arm and 3.12% in the Intervention arm. Thus, the task of discovering the subset of mothers with the highest risk of neonatal deaths involved searching over the discrete/discretized covariates in the BetterBirth study to identify the single subset (stratum) that has the highest rate of neonatal death compared to the global means in both treatment arms i.e., the most anomalous subset.

We found that all mothers with no living children experienced a neonatal death at the time of the study in both treatment arms, i.e., the subset of *mothers with no living children* consisted of 589 participants in the control arm and 642 participants in the treatment arm, all of whom experienced a neonatal death. This finding in the data was further investigated and attributed to a data collection irregularity corresponding to the question in the survey, i.e., how many living children does the pregnant mother have? As a result, this information is discarded from our subsequent analysis. We then further analyzed the vulnerable subset of mothers with neonatal death outcomes after the variable with quality was removed. The finding shows that *low birth-weight* a high-risk factor, which is validated by the domain experts. These two findings demonstrate the power of complementary and collaborative work between domain experts and data scientists that led to the discovery of hidden data quality issues and the validation of domain-expert insights with data-driven results.

Leveraging data sources that are available in silos for more informed insight extraction and decision making

Health surveys, such as the DHS, KI, and PMA, are conducted for different reasons and can be used to help understand several aspects of MNCH[19]. Such a process may involve a range of research questions, such as predicting the likelihood of the outcome, detection of vulnerable groups, and identification of main determinants[5–8,25]. Though these existing surveys have been employed to derive actionable insights, e.g., for devising the new interventions or policies, more could be done to facilitate their effective utilization, especially considering the significant resources (human capital, time, money, etc.) incurred to collect, process, store, and maintain these surveys.

In practice, a critical challenge is the siloed analysis of single survey data. Often there are multiple MNCH-related surveys even in a single country. These surveys might differ in study samples, the information

Table 1 | Our approach identified a data collection irregularity in the BetterBirth study where *mothers with no living children* are found to be extremely risky to neonatal death[9]

	Treatment arms	
	Control	Intervention
Number of mothers in the whole treatment arm	75004	77103
Percentage of neonatal deaths in the treatment arm	3.15%	3.12%
Identified vulnerable group in the treatment arm	Mothers with no living children	Mothers with no living children
Number of identified mothers in the identified group	589	642
Percentage of neonatal deaths in the identified group	100%	100%

Collaborations with the domain experts proved the data collection irregularity for this particular variable and hence it is removed from further analysis.

collected, and the time periods in which the surveys were conducted. Moreover, the variations across health surveys pose a challenge for the integrated use of the different data sources. Variations include but are not limited to specific research questions being addressed using the survey, the entities funding the research project or data collection, the proprietors of the data, and the stipulations and willingness surrounding data access. Furthermore, the extraction of estimates and insights from single surveys is often insufficient, particularly when the surveys suffer from data scarcity challenges such as small sample sizes and data imbalance (i.e., a rare occurrence of an outcome). Thus, despite huge investments and efforts to collect single independent surveys, they often do not meet the increasing information demand for policy- and decision-making. Consequently, there is an established need for combining multiple independent surveys in order to capture distinct characteristics available across these surveys and potentially result in more practical insights.

Next, we illustrate the potential benefits of harmonizing different health surveys related to child mortality. By aggregating these disparate data sets, we can create a more comprehensive and insightful repository of information that can significantly enhance our understanding and ability to improve child mortality outcomes. To this end, we demonstrated a data-driven approach to integrate DHS[16], PMA[18], and KI[17] survey datasets[19]. DHS contains representative data on population, health, HIV, and nutrition through more than 300 surveys in over 90 different countries. These nationally representative surveys are designed to collect data on monitoring and impact evaluation indicators important for individual countries and cross-country comparisons. PMA comprises surveys related to households, service delivery points, and GPS of the area, collected using innovative mobile technology. In addition to the household, individual-level data were collected for each eligible female-identified in the household roster. KI consists of different studies some of which are controlled trials on the child growth effect of different interventions. We particularly used the Alliance for Maternal and Newborn Health Improvement surveys, particularly (AMANHI-1) and (AMANHI-2), which focus on child mortality[19].

Our approach to linking different surveys begins with projecting these surveys into equal-dimensional covariate representations so that samples in these surveys can be compared directly[19]. We employ different techniques to reduce the original covariate dimensions in the surveys to be combined. These techniques include using common covariates among the surveys, dimensionality reduction using principal component analysis, dimensionality reduction using denoising autoencoders, and feature importance rankings. Subsequently, the similarity of samples across disjoint surveys is computed using a distance metric, from which close neighbors are extracted. Next, unique covariates of close neighbors are aggregated and combined with the original study, thereby augmenting the covariate representation of the original survey for better predictive performance[19]. This linking approach is straight forward, and it provides data-level integration of

different surveys, which can minimize resource utilization compared to extra data collection or more sophisticated post-model linkage practices.

We validated our data linkage approach by comparing it against random linkages[19]. First, we separately linked disjoint datasets obtained from DHS data from Burkina Faso, Nigeria, and Ghana. Next, we separately trained models for predicting child mortality using the linked datasets. Lastly, we assessed the performance of the trained models using area under receiver operating characteristic. We found that the models trained on the data linked by our approach significantly outperformed the models train on the siloed datasets or the randomly linked datasets[19]. Interestingly, across different dimension reduction techniques evaluated, using auto-encoders provided the largest improvement, suggesting the potential benefit of recent advances in the machine learning and deep learning domain. Generally, the proposed framework involves the utilization of multiple surveys to: (1) maximise the efficiency of a particular study by incorporating discriminative and unique covariates from another study; (2) improve prediction performance and identify distinctively useful covariates across studies; and (3) provide domain experts and policymakers with additional insights on existing studies and further recommendations for future data collection efforts.

Identifying heterogeneous effects of MNCH interventions

The MNCH domain is characterized by the application of different interventions aimed at reducing preventable maternal and newborn deaths and complications across populations composed of varying characteristics. Improving the quality of care in MNCH, therefore, requires a sound understanding of the varying (heterogeneous) treatment effects of interventions across individuals or subgroups in an MNCH population. This nuanced approach is crucial to activities such as targeted intervention planning in MNCH. Unfortunately, however, intervention impact studies in MNCH predominantly investigate the average treatment effects of interventions across studied populations[23,24]. Often times, this correctly leads to wide acceptance and reuse of interventions proven to be impactful but inadvertently results in the understudying of less effective interventions with limited understanding of the potential reasons related to ineffectiveness. They may also fail to evaluate how well the less-impactful intervention can be expected to work for specific individuals or subgroups of a population[24], or why the intervention did not work for the remaining studied population[9].

Some of the key reasons why the analysis of heterogeneous treatment effects is challenging include the lack of clarity regarding the goals of such analyses and the lack of appropriate approaches to conduct, report, interpret, and apply results from such studies[26]. For example, traditional approaches for analyzing heterogeneous treatment effects typically rely on manual stratification that is often limited to a handful of features selected a priori by domain experts. Fortunately, recent advancements in data-driven subgroup analysis methods enable scalable and unbiased heterogeneous treatment effect analysis[9]. These novel analytic approaches could overcome challenges associated with traditional methods for analyzing subpopulation level effects of interventions in MNCH.

By way of example, after we evaluated the data quality in the BetterBirth study and discarded noisy variables, we proceeded further to uncover potential heterogeneous treatment effects. Recall that the BetterBirth study evaluated the impact of the WHO Safe Childbirth Checklist using a matched-pair, cluster-randomized, controlled trial in 120 government health facilities across 24 districts in Uttar Pradesh India[24]. In this study, the intervention arm population included mother-baby dyads registered for labor and delivery in 60 health facilities that implemented the BetterBirth program. The intervention was the implementation of the WHO Safe Childbirth Checklist, a quality improvement tool that promotes systematic adherence to 28 evidence-based practices associated with improved childbirth outcomes and is primarily used by birth attendants during and after the delivery and before discharge. The control arm was composed of mother-baby dyads registered for labor and delivery in the remaining 60 health facilities that applied the existing standard of care. The primary outcome of interest was a composite outcome of perinatal death, maternal

death, or severe maternal complications occurring within the first 7 days after delivery. The study enrolled and determined the outcomes of over 157,000 eligible participants across the intervention and comparison groups. The study concluded that although adherence to good birth practices was higher in the intervention arm, maternal mortality, perinatal mortality, and maternal morbidity did not differ significantly between the treatment and control arms[24].

One can pose a few critical questions based on the BetterBirth study findings:

Q_1: *though the intervention did not significantly reduce the primary outcomes in the intervention arm (compared to the control arm), was there a subset of mother-baby dyads who actually benefited from the intervention?*

Q_2: *if such a subset of mother-baby dyads that benefited from the intervention exist, what are the characteristics of this subset?*

We tried to answer these questions by looking for potential heterogeneous treatment effects in the BetterBirth study[9]. Procedurally, we first trained a logistic regression model on the control arm data to predict the likelihood of developing the binary primary composite outcome. Even though logistic regression was employed due to its simplicity and ease of interpretation, other classification algorithms, such as Gradient Boosting, could also be used for this task. Next, we use the trained classification model to estimate the expected outcome for each mother-baby dyads records in the intervention group. Finally, we applied subset scanning from the anomalous pattern detection literature[27] to identify the subset of mother-baby dyads in the intervention arm that had the largest deviation between the actual outcomes and expected outcomes. We found that mother-baby dyads described by *normal gestational age at birth*, *known parity*, and *unknown number of abortions* were found to benefit from the Checklist intervention significantly (Odds Ratio: 0.70, 95%S Confidence Interval: 0.62–0.79, with empirical *p*-value < 0.001). However, it is worth noting that such insights are still hypothetical and confirmatory studies e.g., through adaptive randomization[28] are critical to verify such generated hypotheses.

Tracking temporal data distribution changes in MNCH trends

Maternal, newborn, and child health is influenced by the ever-changing demographics of the population due to factors such as interventions, climate change, pandemics, and civil wars[29–32]. Therefore, it is crucial to recognize and analyze these changes across various regions and administrative units over time, as well as the prevalence of outcome changes over time. With this in mind, our objective is to highlight the limitations of country-level aggregated or averaged reports of outcomes, such as under-5 child mortality, which may not provide a comprehensive view of the situation. For example, country-level reports do not reflect on regions or sub-populations that are still lagging behind or the regions or sub-populations that are faring better than the reported average. These aggregated reports can often obscure the realities faced by subsets of the population that fall on the extreme ends of the country-level average. By delving deeper and exploring these subsets, we can gain a clearer understanding of the true scope and impact of maternal, newborn, and child health challenges.

Our previous study demonstrated how temporal data distribution changes, such as the concept drift in the statistical properties of a child mortality across time[33], which can also be used to investigate variability in further outcomes in MNCH. We leveraged data-driven subgroup discovery to identify the sub-populations of women that experience larger than expected changes in under-5 mortality rates between two points in time, approximately 10–15 years apart[33]. Procedurally, we begin by training and calibrating a machine learning predictive model to predict the likelihood of under-5 mortality using nationally representative DHS[16] data from an earlier time-point (T_0). Second, we apply the predictive model to predict the probability of under-5 mortality at a more recent time-point (T_1). Third, we compute the change in the odds of the outcome between the two time steps T_0 and T_1. Lastly, we apply subset scanning[27] to identify the sub-populations

in the T_1 data whose outcomes differ the most from their predicted probabilities based on the T_0 model.

By applying this approach, we found several potentially interesting findings. For example, this approach suggests that in Ethiopia, households composed of single mothers with 2 children reported the largest decrease in under-5 mortality, i.e., from 47% (in 2000) to less than 7.5% (in 2016). Similarly, in Nigeria, households residing in the South or South-West regions experienced the largest decrease in under-5 mortality (from 14.8% to 7.4%). Further work is still needed to study causal connections related to observed sub-population-level changes in under-5 mortality trends.

Improving the interpretability of "black box" machine learning models for MNCH domain experts

Whereas most ML models are good at prediction and classification, they are often not readily trusted and adopted by MNCH stakeholders due to their "black box" nature. For ideal use in decision-making and intervention planning, MNCH stakeholders and policymakers require models that are not only accurate but also interpretable and able to generate actionable insights. Consequently, machine learning practitioners and adopters must develop and use methods for inspecting "black box" models to generate actionable insights and improve the trustworthiness of their proposed solutions.

By way of example, we conducted a study that aimed at identifying the factors associated with neonatal mortality by analyzing the DHS[16] survey datasets from 10 Sub-Saharan countries[25]. For each survey dataset, we trained an ensemble gradient boosting classifier that was used to identify mothers who experienced a neonatal death within 5 years prior to participating in the survey. To improve explainability and identify new insights, we visualized the feature importance and partial dependence of features in the model. Herein, feature importance refers to the ranked list of the most important features contributing to the prediction in the ensemble model. Partial dependence refers to the relationship between a single feature and an outcome of interest, holding other features constant, i.e., how, on average, changing the values of a given feature while holding the values of all other features affects the risk of a given outcome.

Interestingly, through these "black box" model inspection techniques, we confirmed the positive correlation between birth spacing and risk of neonatal mortality and identified a plausible negative correlation between household size and risk of neonatal mortality[25]. We also established that mothers living in smaller households have a higher risk of neonatal mortality than mothers living in larger households.

Discussions and future directions

MNCH in low-and middle-income settings, often in the Global South, has been the primary focus of a number of United Nation's Sustainable Development Goals (SDGs), particularly SDG-3 - *Ensure healthy lives and promote well-being for all at all ages*. Though encouraging improvements have been reported over recent years[4], partly due to a number of successful interventions, many countries are still lagging behind the SDG targets pertaining to MNCH. Challenges, such as COVID-19 pandemic, climate change, natural disasters and civil wars[29–32], further complicate the current MNCH situation, e.g., with adverse impacts on health facilities. Fortunately, we have witnessed an astonishing rise of data-driven capabilities, such as machine learning, over recent years that could be employed to help understand MNCH challenges from a plethora of different data sources. Domain experts, such as public health professionals, are often at the forefront of studies conducted in the MNCH domain due to their accumulated knowledge facilitated by long-term on-ground experiences. On the other hand, recent advances in data-driven approaches demonstrate capabilities for analyzing data and extracting actionable insights for domain experts in an efficient and scalable way[9,25,33]. Thus, it is critical for these two communities to collaborate and complement each other in a bid to solve the MNCH challenges and accelerate the progress toward achieving SDG targets related to MNCH. Furthermore, an intersectional approach, which brings different stakeholders and their perspectives, is key to better understanding the fundamental MNCH challenges. The intersectional approach should

encompass all the steps, starting from problem formulation, data collection design, and collection to analysis of the data collected and driving actionable insights. The stakeholder group may involve community health workers, mid-level and national-level health system administrators, non-profit organizations, and policymakers. The data-driven approach has a tremendous opportunity to facilitate communications among these diverse groups by providing insights that are intrinsic and understandable to these groups.

Particularly, we foresee wider adoptions of similar data-driven technologies that aim to utilize multitude of MNCH data sources that are often collected and used in silos across different MNCH challenges. In addition to DHS[16], KI[17], and PMA[18], there are other MNCH data sources that could be utilized in future data-driven solutions for MNCH. Examples include the Multiple Indicator Cluster Survey (MICS)[34], which provide a wide range of indicators including those on the health, nutritional status, and education of children and women. Moreover, MICS surveys have been collected in subsequent rounds (https://mics.unicef.org/) providing more frequently collected data in cost effective manner, which makes it suitable for longitudinal studies, e.g., tracking SDGs. District Health Information Software (DHIS2)[35] is another data source that could be considered for similar tasks, e.g., the malnutrition data being collected at the health facility level in Kenya could be utilized to further forecast acute malnutrition hot-spots in the future. In addition, complementary data sources, other than health surveys, need to be evaluated and used in data-driven approaches to further strengthening the understanding of machine learning models towards complex problems. These complimentary data sources include satellite imagery, which are often freely available from multiple providers. These remotely sensed images provide recent changes on the ground (e.g., expansion of population settlement) and to understand the impact of climate change and disasters (e.g., flood and drought)[36,37].

Recently, Large Language Models (LLMs), which are specific type of models for natural language processing, are demonstrated to possess a higher degree of capability, e.g., as conversational agents[38]. Similar technologies could also be used to democratize access to technologies, e.g., by providing personalized nutrition recommendations during pregnancy[39]. However, there are potential risks that could be associated with these technologies, e.g., hallucination of nonfactual information and misinformation. Thus, we argue that such systems need to demonstrate a level of trustworthiness, e.g., fairness across various segments of the population, reliability and safety, explainability, and protection privacy[10,14]. Moreover, as with any technology, the regulations of such systems are critical to have a standardized adoption of these approaches across borders.

Conclusions

In this perspective, we aim to encourage collaborations among experts from different domains by sharing a diverse set of our prior works on data science and machine learning for MNCH that involved successful collaborations with MNCH experts. Specifically, we highlighted how data-driven techniques shed light on a number of MNCH challenges, such as data-quality, health surveys that are often available in silos, heterogeneous treatment effects, understanding spatio-temporal data distribution shifts, and adding up to the explainability of MNCH models via inspection of machine learning models, which are often treated as 'black box'.

Evidence-based policy-making and intervention designs can strongly benefit from similar data-driven techniques discussed in this paper. However, assessing the quality of data available is a priority before the data is used for decision-making, particularly for MNCH surveys that are prone to a number of quality issues due to the nature of the collection process and/or the number of personals involved in the process. Furthermore, MNCH domain is mostly characterized by the availability of multiple data sources, such as DHS[16], PMA[18], KI[17], MICS[34], DHIS2[35], but in silos with no efficient utilization of their aggregated form. Thus, we demonstrated how aggregation of these data sources, by linking records, helped to improve predictive capabilities of child mortality in a number of Sub-Saharan African countries. While interventions played a significant role in reducing deaths and complications related to mothers, children, and newborns, there is a significant gap in the

literature to investigate interventions that are less impactful in the average population but might have benefited a particular segment of the population.

To this end, we shared that our data-driven methods could automatically identify and characterize sub-populations that have significantly benefited from the interventions using the BetterBirth study as a use case. Similarly, the country-aggregated reports often shared to reflect the improvement of MNCH (e.g., reduction of child mortality rate in a country) are limited to show the whole picture. For example, the regions or sub-populations that lag behind the reported-average are not well studied. To this end, we highlighted how those sub-populations with worse than average child mortality rates are identified in the DHS data by detecting spatio-temporal data distribution changes. Additionally, the perception of machine learning models as 'black boxes' is one of the reasons that restrict the wider adoption of machine learning models by the MNCH domain experts and policy makers. Thus, we have highlighted our work on adding an extra layer of explainability by investigating models designed to predict neonatal mortality. Note that similar approaches described in this perspective could be employed in other domains that are extensions to MNCH. For example, we collaborated with domain experts in family planning and contraception, to extract insights about contraceptive use from the DHS surveys, such as discriminating contraceptive use patterns under different discontinuation reasons, contraceptive uptake distributions, and transition information across contraceptive types[40,41].

References

1. Spector, J. M. & Agrawal, P. et al. Improving quality of care for maternal and newborn health: prospective pilot study of the WHO Safe Childbirth Checklist program. *PLoS ONE* **7**, e35151 (2012).
2. Health, T. L. G. Progressing the investment case in maternal and child health (2021).
3. Hug, L., Sharrow, D. & You, D. Levels and trends in child mortality: report 2017. Tech. Rep. The World Bank (2017).
4. Rahman, M. M. et al. Reproductive, maternal, newborn, and child health intervention coverage in 70 low-income and middle-income countries, 2000–30: trends, projections, and inequities. *Lancet Global Health* **11**, e1531–e1543 (2023).
5. Rutstein, S. O. Factors associated with trends in infant and child mortality in developing countries during the 1990s. *Bull. World Health Organization* **78**, 1256–1270 (2000).
6. Wang, L. Determinants of child mortality in LDCs: empirical findings from Demographic and Health Surveys. *Health Policy* **65**, 277–299 (2003).
7. Hanmer, L., Lensink, R. & White, H. Infant and child mortality in developing countries: analysing the data for robust determinants. *J. Dev. Stud.* **40**, 101–118 (2003).
8. Boschi-Pinto, C., Velebit, L. & Shibuya, K. Estimating child mortality due to diarrhoea in developing countries. *Bull. World Health Organization* **86**, 710–717 (2008).
9. Tadesse, G. A. et al. Principled subpopulation analysis of the BetterBirth study and the impact of WHO's Safe Childbirth Checklist intervention. In *AMIA Annual Symposium Proceedings*, vol. 2022, 1042 (2022).
10. Speakman, S. et al. Detecting systematic deviations in data and models. *Computer* **56**, 82–92 (2023).
11. Polyzotis, N., Zinkevich, M., Roy, S., Breck, E. & Whang, S. Data validation for machine learning. *Proc. Mach. Learn. Syst.* **1**, 334–347 (2019).
12. Budach, L. et al. The effects of data quality on machine learning performance. *arXiv preprint arXiv:2207.14529* (2022).
13. Roberts, M. et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat. Mach. Intell.* **3**, 199–217 (2021).

14. Liang, W. et al. Advances, challenges and opportunities in creating data for trustworthy AI. *Nat. Mach. Intell.* **4**, 669–677 (2022).

15. Harkare, H. V., Corsi, D. J., Kim, R., Vollmer, S. & Subramanian, S. The impact of improved data quality on the prevalence estimates of anthropometric measures using DHS datasets in India. *Sci. Rep.* **11**, 10671 (2021).

16. ICF International: Demographic and Health Surveys (DHS). (Funded by USAID. Rockville, Maryland, 2004–2017).

17. Data Store Explorer: Knowledge Integration (KI) - Africa. http://africa.studyexplorer.io/. Last accessed on March 27, (2024).

18. Performance Monitoring and Accountability 2020 (PMA2020) Project. (Bill & Melinda Gates Institute for Population and Reproductive Health, Johns Hopkins Bloomberg School of Public Health.).

19. Tadesse, G. A. et al. Data-level linkage of multiple surveys for improved understanding of global health challenges. *AMIA Summits on Transl. Sci. Proc.* **2021**, 92 (2021).

20. Kitson, N. K. & Constantinou, A. C. Learning Bayesian networks from demographic and health survey data. *J. Biomed. Inf.* **113**, 103588 (2021).

21. Oala, L. et al. DMLR: Data-centric Machine Learning Research–past, present and future. *arXiv preprint arXiv:2311.13028* (2023).

22. Spector, J. M. & Lashoher, A. et al. Designing the WHO Safe Childbirth Checklist program to improve quality of care at childbirth. *Int. J. Gynecol. Obstetrics* **122**, 164–168 (2013).

23. Delaney, M. M. & Miller, K. A. et al. Unpacking the null: a post-hoc analysis of a cluster-randomised controlled trial of the WHO Safe Childbirth Checklistin Uttar Pradesh, India (BetterBirth). *The Lancet Global Health* **7**, e1088–e1096 (2019).

24. Semrau, K. E. & Hirschhorn, L. R. et al. Outcomes of a coaching-based WHO Safe Childbirth Checklist program in India. *New England J. Med.* **377**, 2313–2324 (2017).

25. Ogallo, W. et al. Identifying factors associated with neonatal mortality in Sub-Saharan Africa using machine learning. In *AMIA Annual Symposium Proceedings*, vol. 2020, 963 (2020).

26. Varadhan, R., Segal, J. B., Boyd, C. M., Wu, A. W. & Weiss, C. O. A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. *J. Clin. Epidemiol.* **66**, 818–825 (2013).

27. Cintas, C. et al. Pattern detection in the activation space for identifying synthesized content. *Pattern Recognition Letters* **153**, 207–213 (2022).

28. Russell, D., Hoare, Z., Whitaker, R., Whitaker, C. & Russell, I. Generalized method for adaptive randomization in clinical trials. *Stat. Med.* **30**, 922–934 (2011).

29. Luyten, A., Winkler, M. S., Ammann, P. & Dietler, D. Health impact studies of climate change adaptation and mitigation measures–a scoping review. *J. Clim. Change Health* **9**, 100186 (2023).

30. Ahmed, T. et al. The effect of COVID-19 on maternal newborn and child health (MNCH) services in Bangladesh, Nigeria and South Africa: call for a contextualised pandemic response in LMICs. *Int. J. Equity Health* **20**, 1–6 (2021).

31. Akseer, N. et al. Women, children and adolescents in conflict countries: an assessment of inequalities in intervention coverage and survival. *BMJ Global Health* **5**, e002214 (2020).

32. Jawad, M., Hone, T., Vamos, E. P., Cetorelli, V. & Millett, C. Implications of armed conflict for maternal and child health: a regression analysis of data from 181 countries for 2000–2019. *PLoS Med.* **18**, e1003810 (2021).

33. Idrees, I., Speakman, S., Ogallo, W. & Akinwande, V. Successes and misses of global health development: detecting temporal concept drift of under-5 mortality prediction models with bias scan. *AMIA Summits Transl. Sci. Proc.* **2021**, 286 (2021).

34. Khan, S. & Hancioglu, A. Multiple indicator cluster surveys: delivering robust data on children and women across the globe. *Stud. Family Planning* **50**, 279–286 (2019).

35. Dehnavieh, R. et al. The District Health Information System (DHIS2): a literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. *Health Inf. Manag. J.* **48**, 62–75 (2019).

36. Johnson, K. B., Jacob, A. & Brown, M. E. Forest cover associated with improved child health and nutrition: evidence from the Malawi Demographic and Health Survey and satellite data. *Glob. Health: Sci. Pract.* **1**, 237–248 (2013).

37. Curto, A. et al. Associations between landscape fires and child morbidity in Southern Mozambique: a time-series study. *Lancet Planetary Health* **8**, e41–e50 (2024).

38. Biswas, S. S. Role of ChatGPT in public health. *Ann. Biomed. Eng.* **51**, 868–869 (2023).

39. Tsai, C.-H. et al. Generating personalized pregnancy nutrition recommendations with GPT-Powered AI chatbot. In *Proceedings of the 20th International Conference on Information Systems for Crisis Response and Management* (2023).

40. Cintas, C. et al. Data-driven sequential uptake pattern discovery for family planning studies. In *AMIA Annual Symposium Proceedings*, vol. 2021, 324 (2021).

41. Cintas, C. et al. Decision platform for pattern discovery and causal effect estimation in contraceptive discontinuation. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 5288–5290 (2021).

Acknowledgements

Author contributions
G.A.T., W.O., C.C., S.S., A.W.-B., and C.W. conceived the manuscript and designed the structure. G.A.T., W.O., and C.C. helped draft the manuscript. All authors helped revise the manuscript. G.A.T., W.O., and C.C. helped implement the comments raised during revisions. All authors approved the submitted and revised versions.

Competing interests
The authors declare no competing interests.

Additional information